A quality assurance process of a surface wind database in Eastern Canada

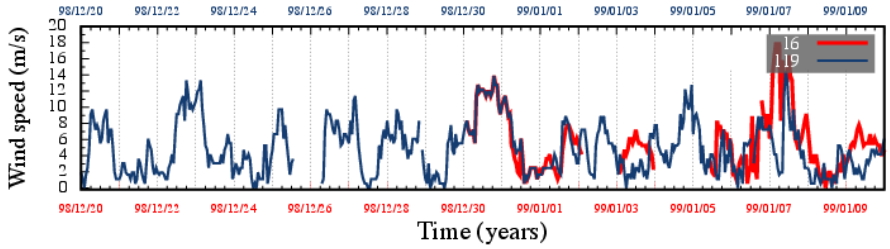
E.E. Lucio-Eceiza	**UCM, IGEO**
J.F. González-Rouco	**UCM, IGEO**
J. Navarro	**CIEMAT**
A. Hidalgo	**Global Forecasters S.L.**
P.A. Jiménez	**CIEMAT**
E. García-Bustamante	**U. Giessen**
J.L. Conte	**Global Forecasters S.L.**
H. Beltrami	**CASI**

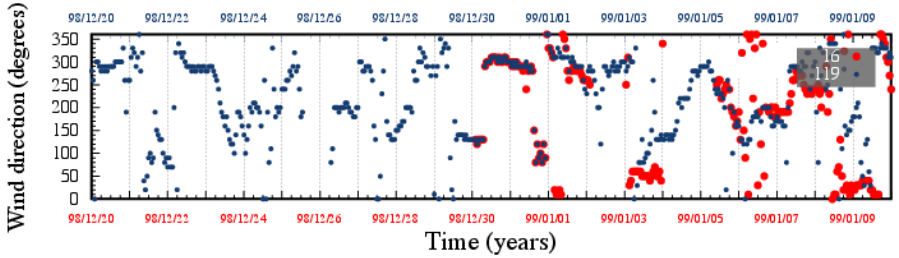
Paleoclimate Modeling & Analysis (PalMA)
Dpto. Física de la Tierra, Astronomía y Astrofísica II
Universidad Complutense de Madrid
Instituto de Geociencias CSIC-UCM

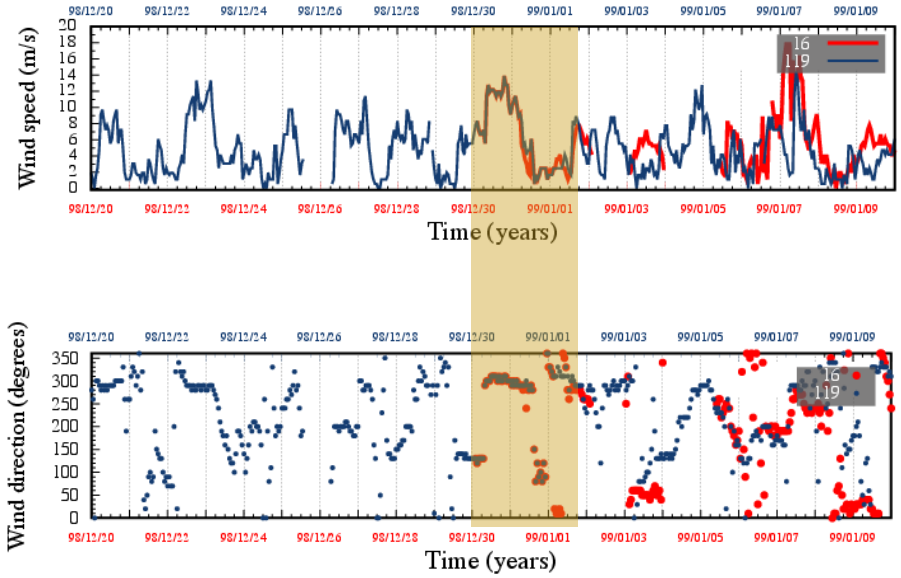
Motivation

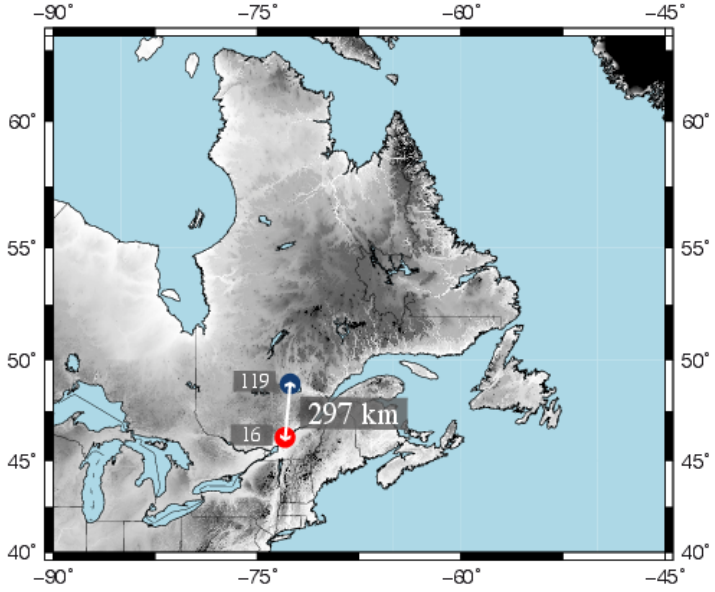
Database Compilation

Quality Assurance









References:

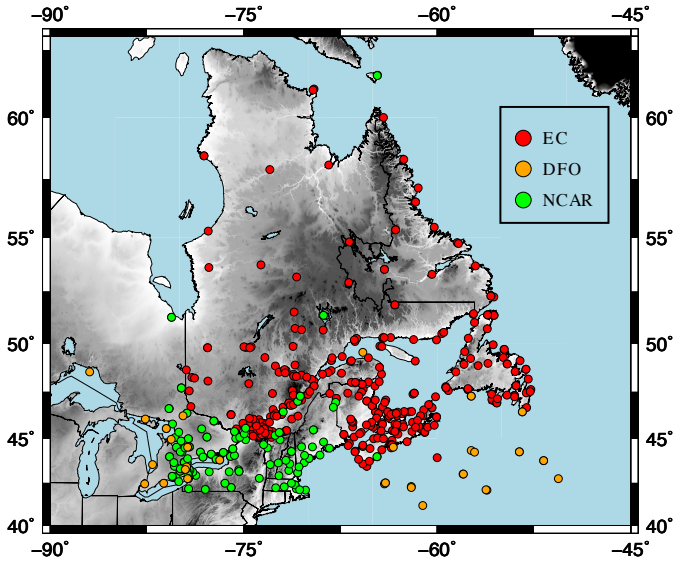
- García-Bustamante E. PhD Thesis: “Statistical Downscaling of surface wind field and wind energy over a complex terrain area”. Supervisors: J. F. González-Rouco y J. Navarro. Universidad Complutense de Madrid. 2010.

- Jiménez, P. A. PhD Thesis: "Analysis of surface wind over complex terrain: a dynamical downscaling study with the WRF model". Supervisors: J. F. González-Rouco y J. Navarro. Universidad Complutense de Madrid. 2009.

Motivation

Database Complitaion

Quality Assurance



Nova Scotia

Newfoundland & Labrador

New Brunswick

Environment Canada (EC)

Prince Edward Island

Quebec

Great Lakes

Canadian East Coast

Department of Fisheries & Oceans (DFO) (buoys)

Vermont

New York

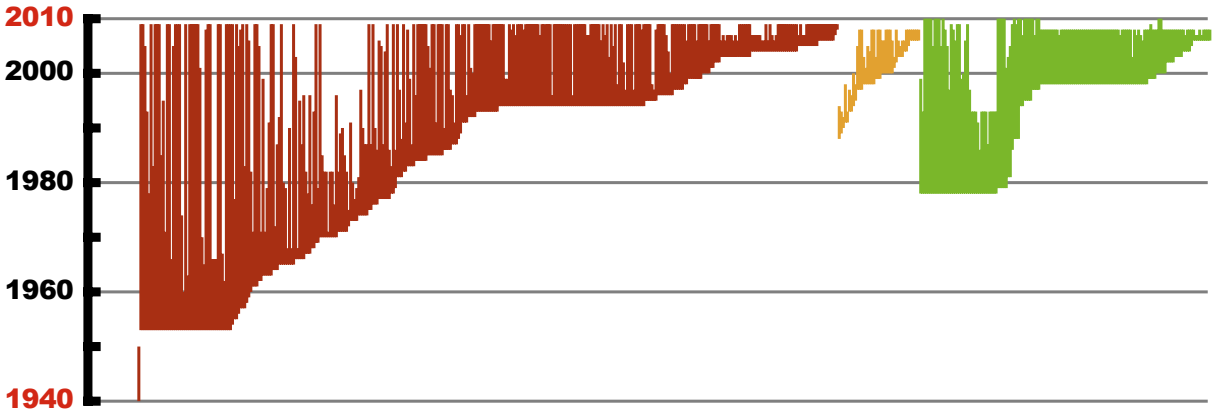
Maine

National Center for Atmospheric Research (NCAR)

New Hampshire

Massachusetts

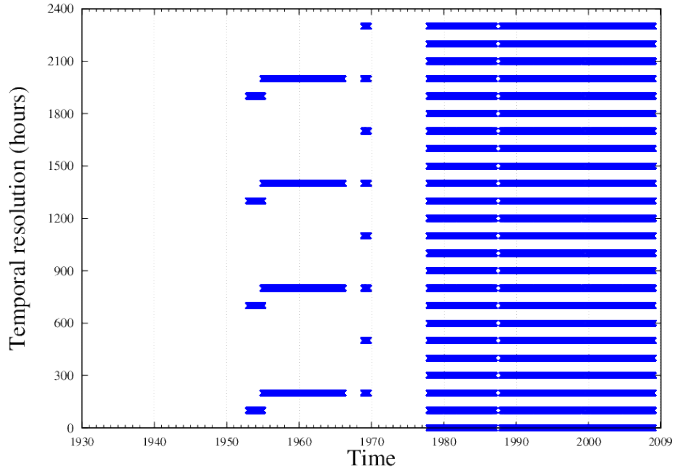
Lifetime of the stations



Initial and final year of operation for each station (red for EC, orange for DFO, green for NCAR). Each station is a vertical line.

The database:

- 527 stations: 344 from EC, 40 buoys and 143 from NCAR.
- Time resolution: hourly, 3hourly and synoptic, sometimes within the same station.
- Time span: beginning at 1940 ending at 2010.

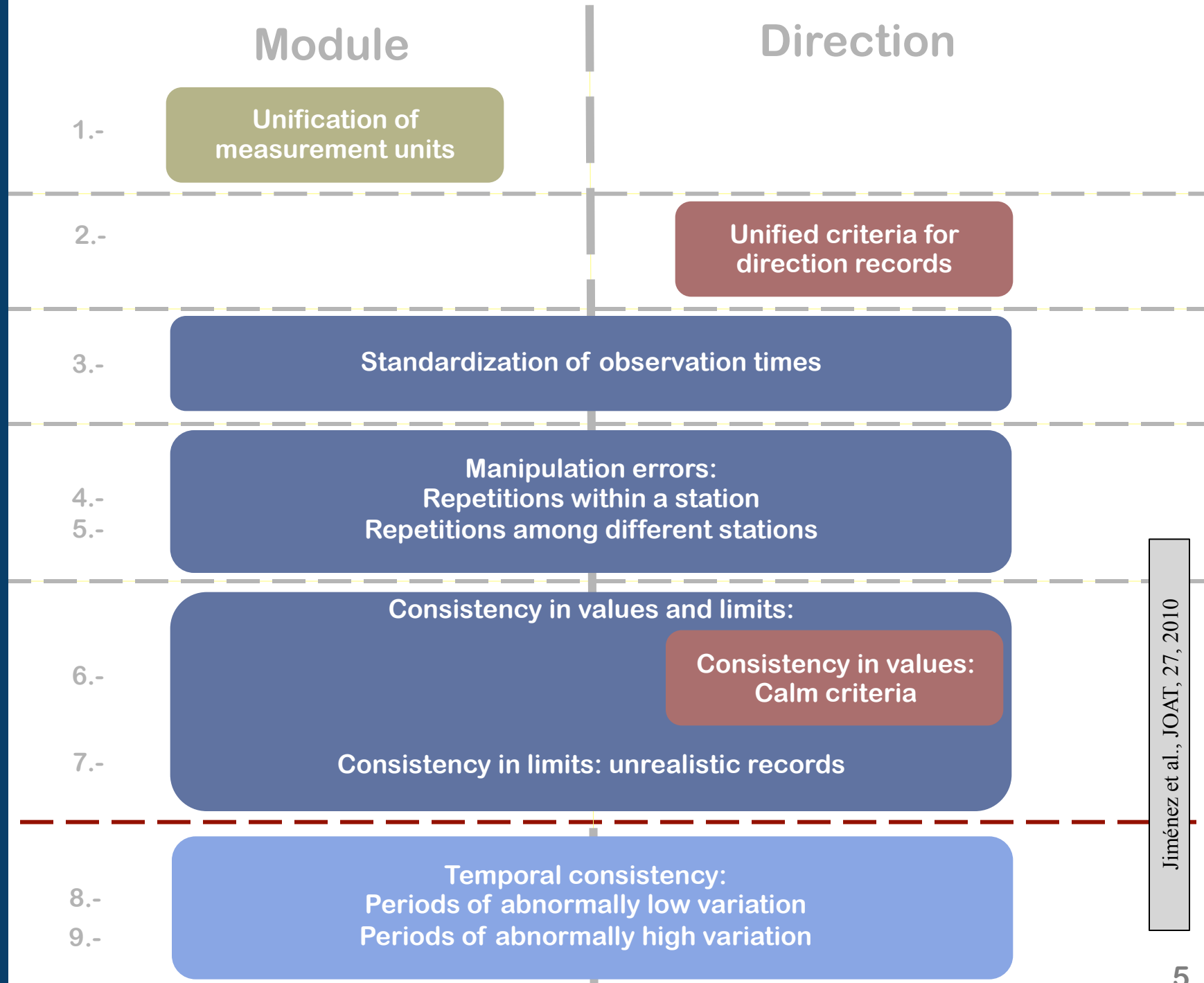


Example: station 291 (EC), evolves from synoptic to 3hourly time-resolution, the late decades are in hourly.

Motivation

Database Compilation

Quality Assurance



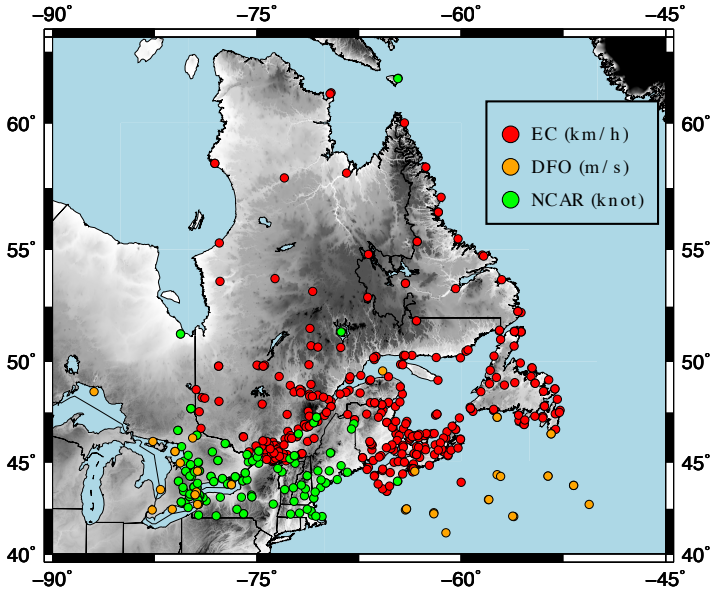
Jiménez et al., JOAT, 27, 2010

Unification of measurement units

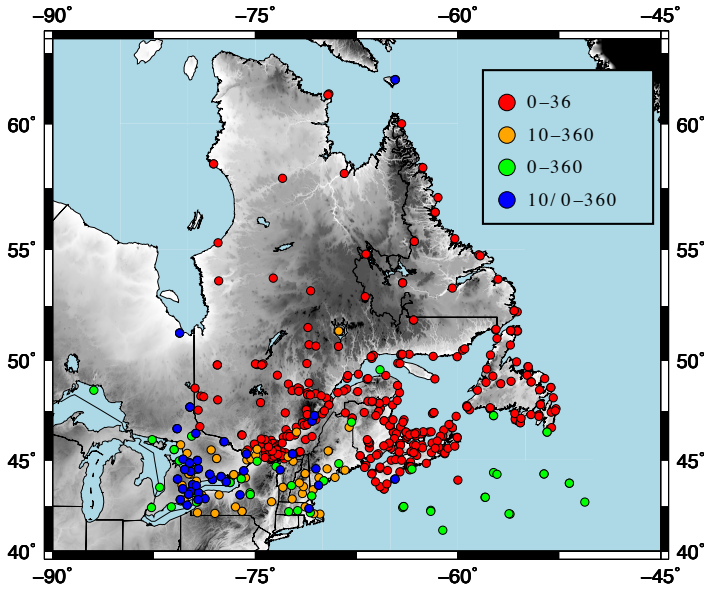
All measurements must be in m/s.

Unified criteria for direction records

All measurements must be in units of degrees.



Spatial distribution of sites corresponding to different institutions and their measurement units.

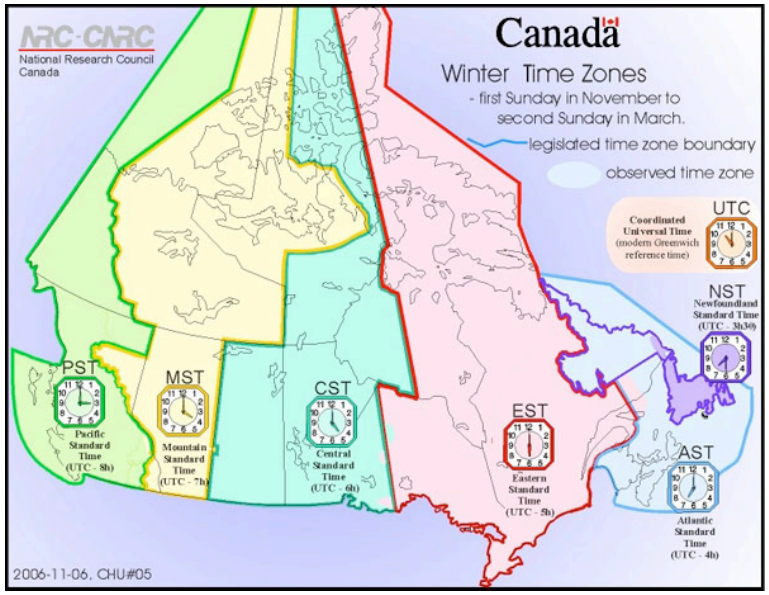


Spatial distribution of wind direction ranges.

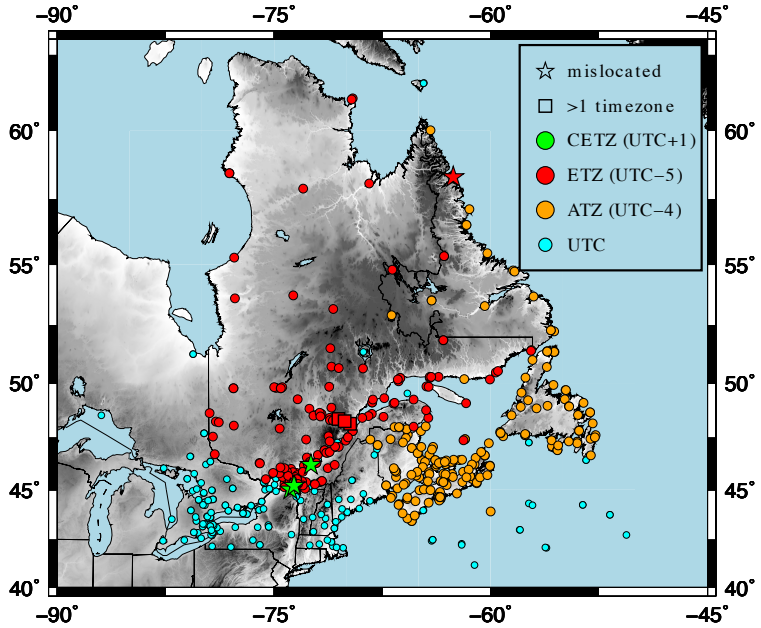
Motivation
Database Compilation
Quality Assurance

	Module	Direction
1.-	Unification of measurement units	
2.-		Unified criteria for direction records
3.-	Standardization of observation times	
4.-	Manipulation errors: Repetitions within a station	
5.-	Repetitions among different stations	
6.-	Consistency in values and limits:	Consistency in values: Calm criteria
7.-	Consistency in limits: unrealistic records	
...		

- **DFO & NCAR record in UTC.**
- **EC does it in Local Standard Time (LST).**
- **All stations are transformed to UTC.**



Time Zones distribution over Canada.
Source: http://www.nrc-cnrc.gc.ca/eng/services/inms/time-services/time-zones.html



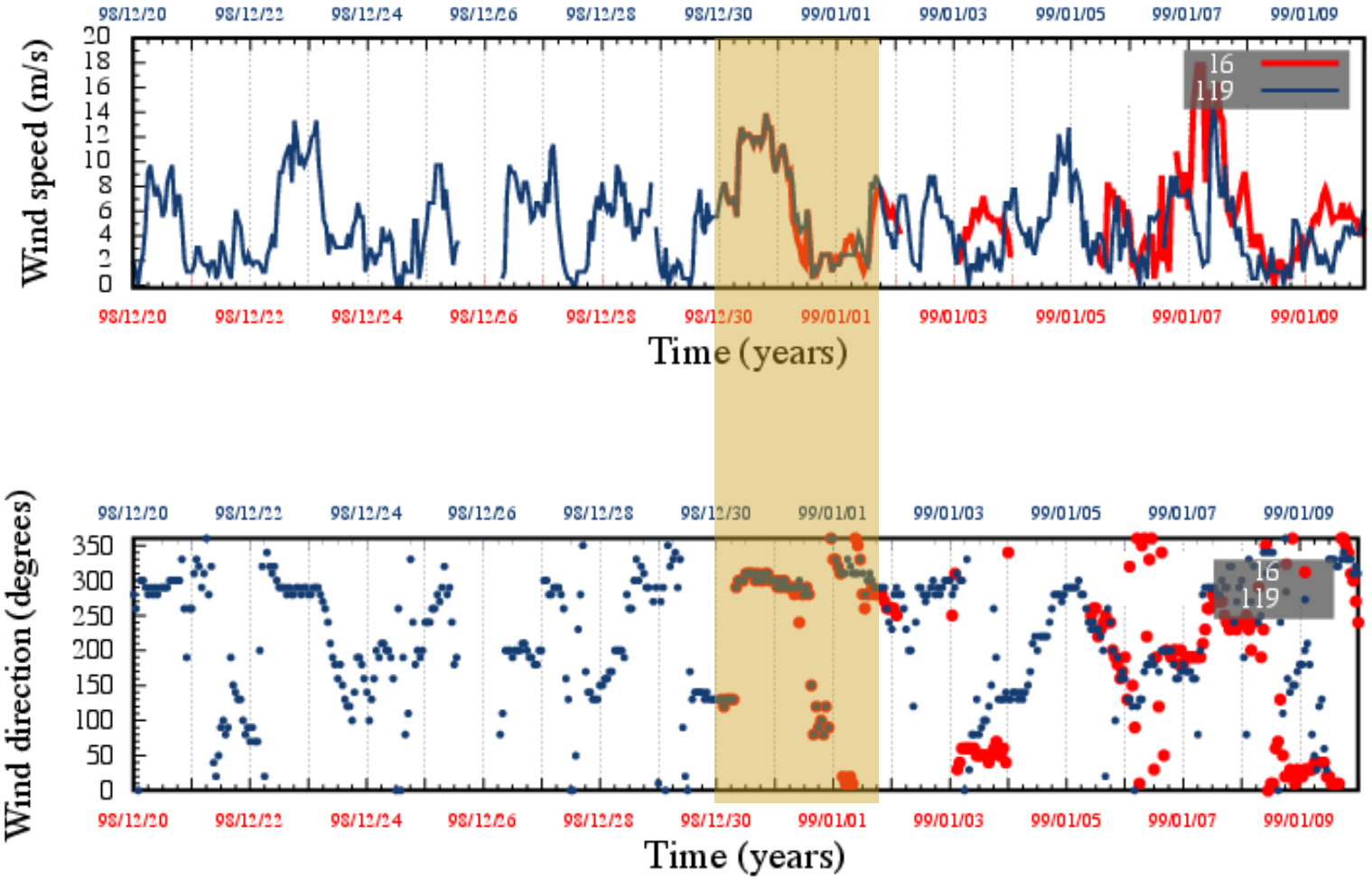
Spatial distribution of the time zones that the staions belong to, three are located in Central European Time Zone (CETZ) despite belonging to Quebec.

Motivation

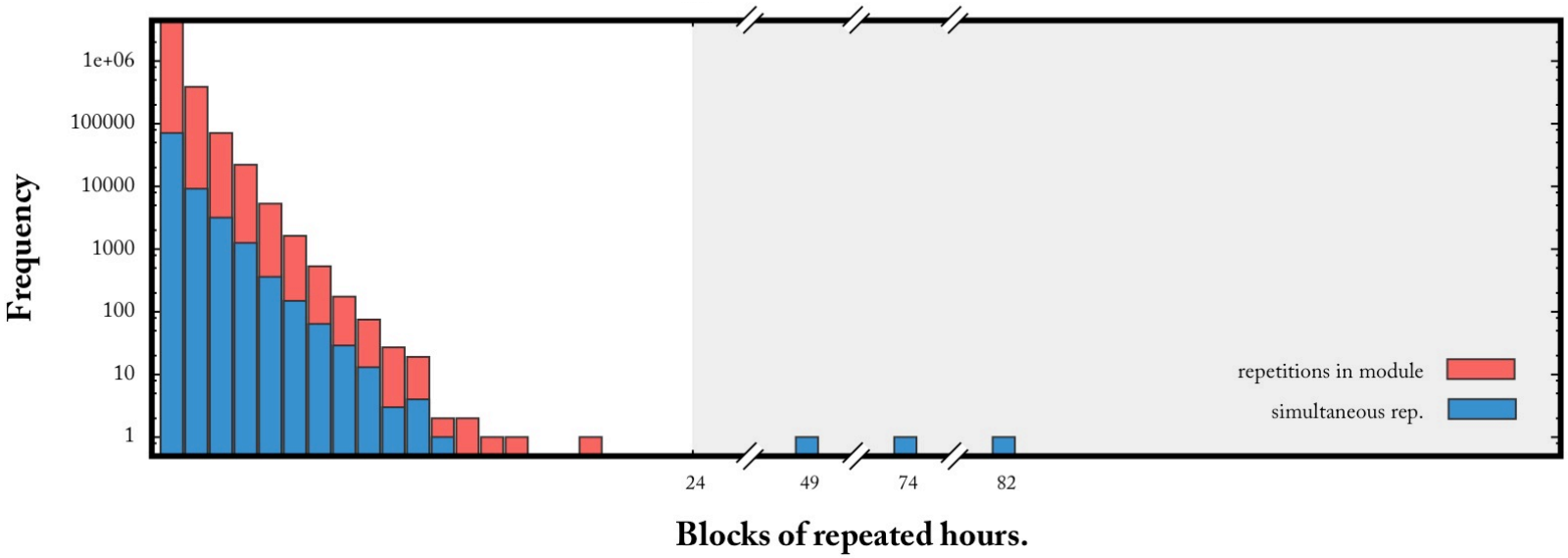
Database Compilation

Quality Assurance

	Module	Direction
1.-	Unification of measurement units	
2.-		Unified criteria for direction records
3.-	Standardization of observation times	
4.-	Manipulation errors: Repetitions within a station	
5.-	Repetitions among different stations	
	Consistency in values and limits:	
6.-		Consistency in values: Calm criteria
7.-	Consistency in limits: unrealistic records	
...		

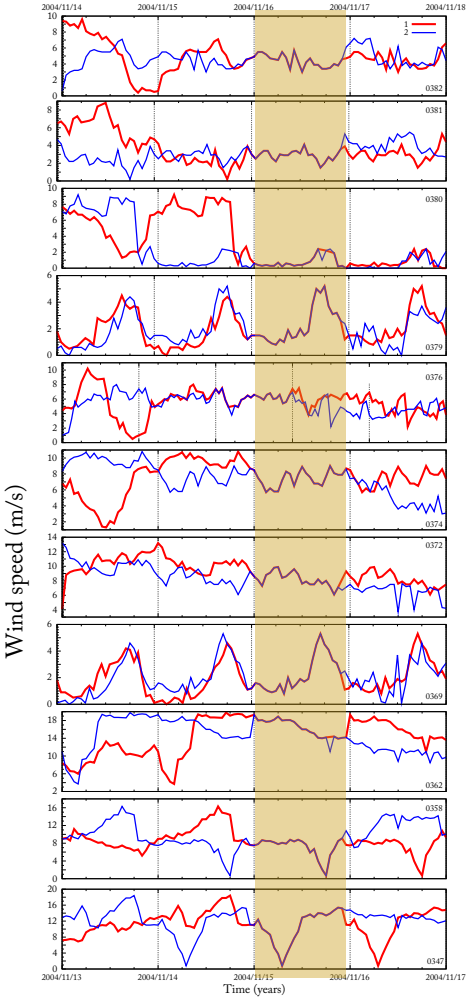


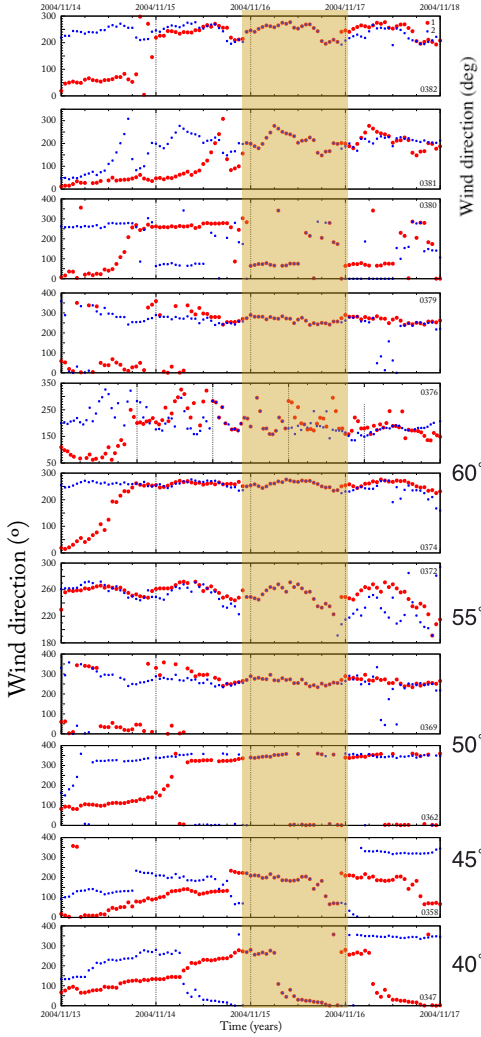
Stations 16 and 119, both belonging to EC that share 40 hours of data, despite being ~300 km apart from each other.



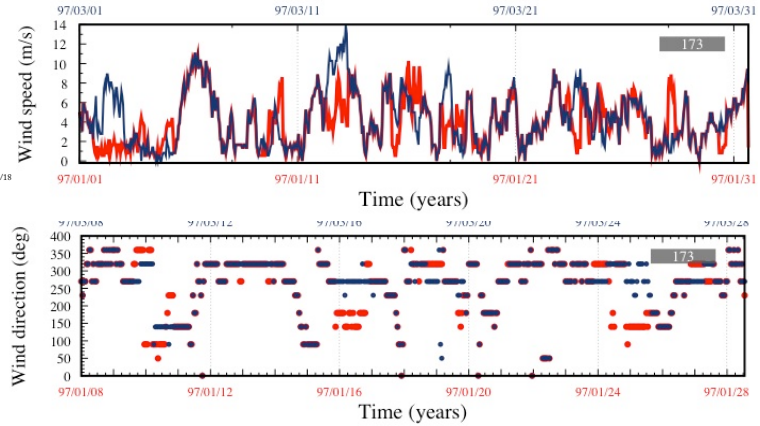
Frequency distribution for the repetitions within a station: in red the repetitions for module and in blue the repetitions that take place simultaneously both in module and direction.

Repetitions within a station:

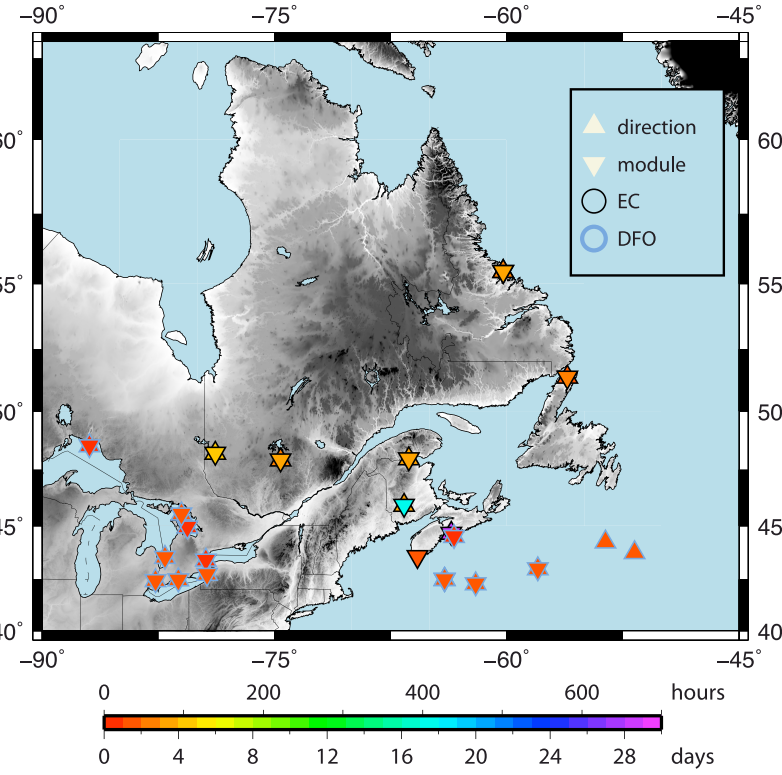




11 buoys simultanously had a failure that copied one day in the next one, module (left) and direction (right).

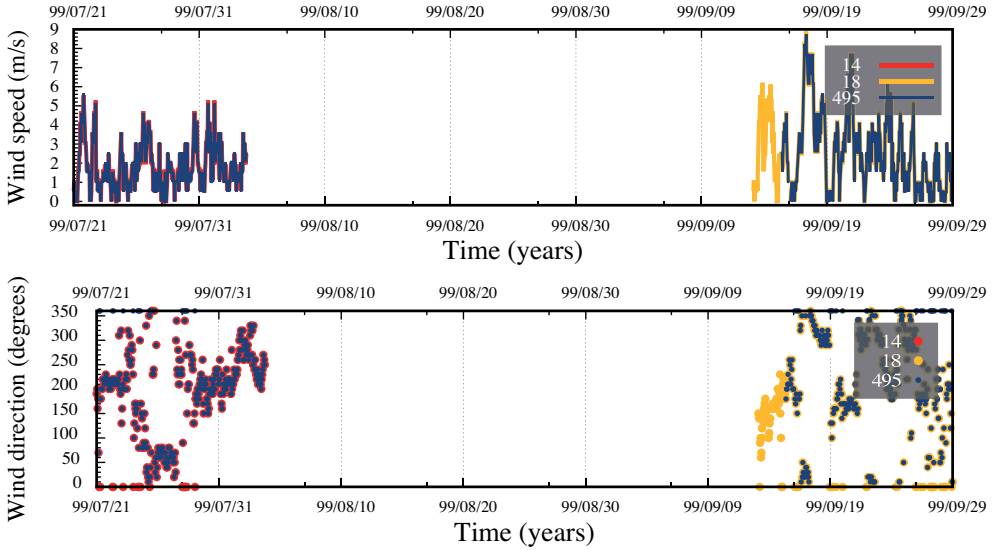


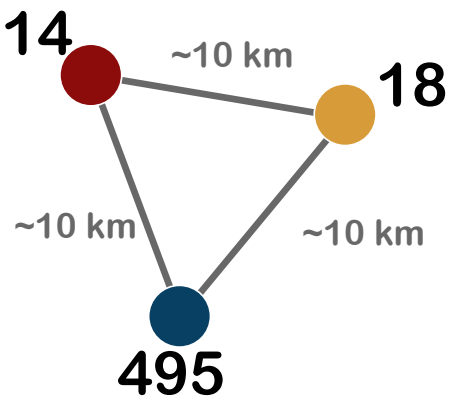
Station that shares almost a month (but not entirely).



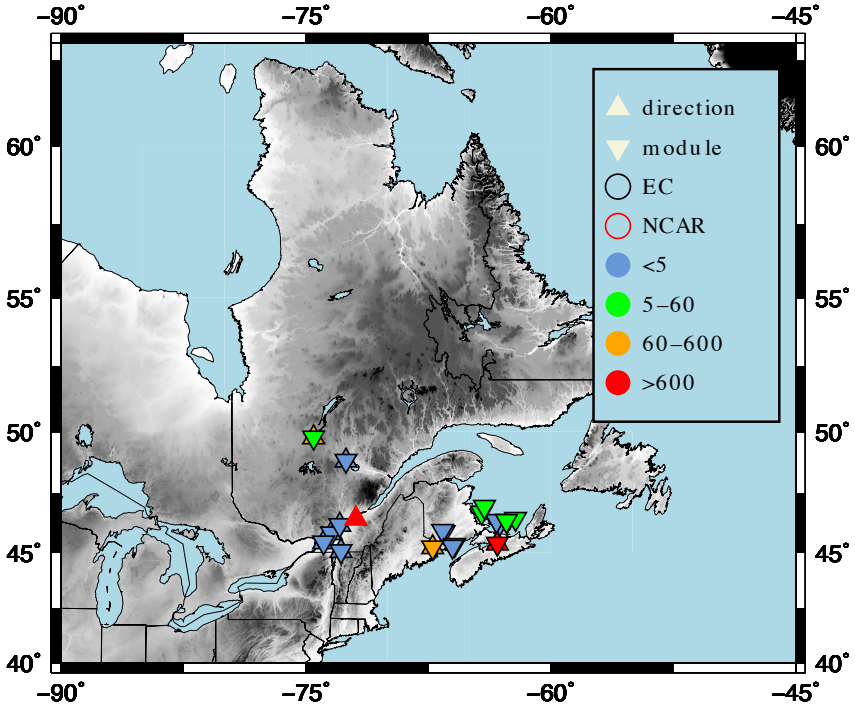
Number of days found to be repeated at various sites in the database.

Repetitions among stations:



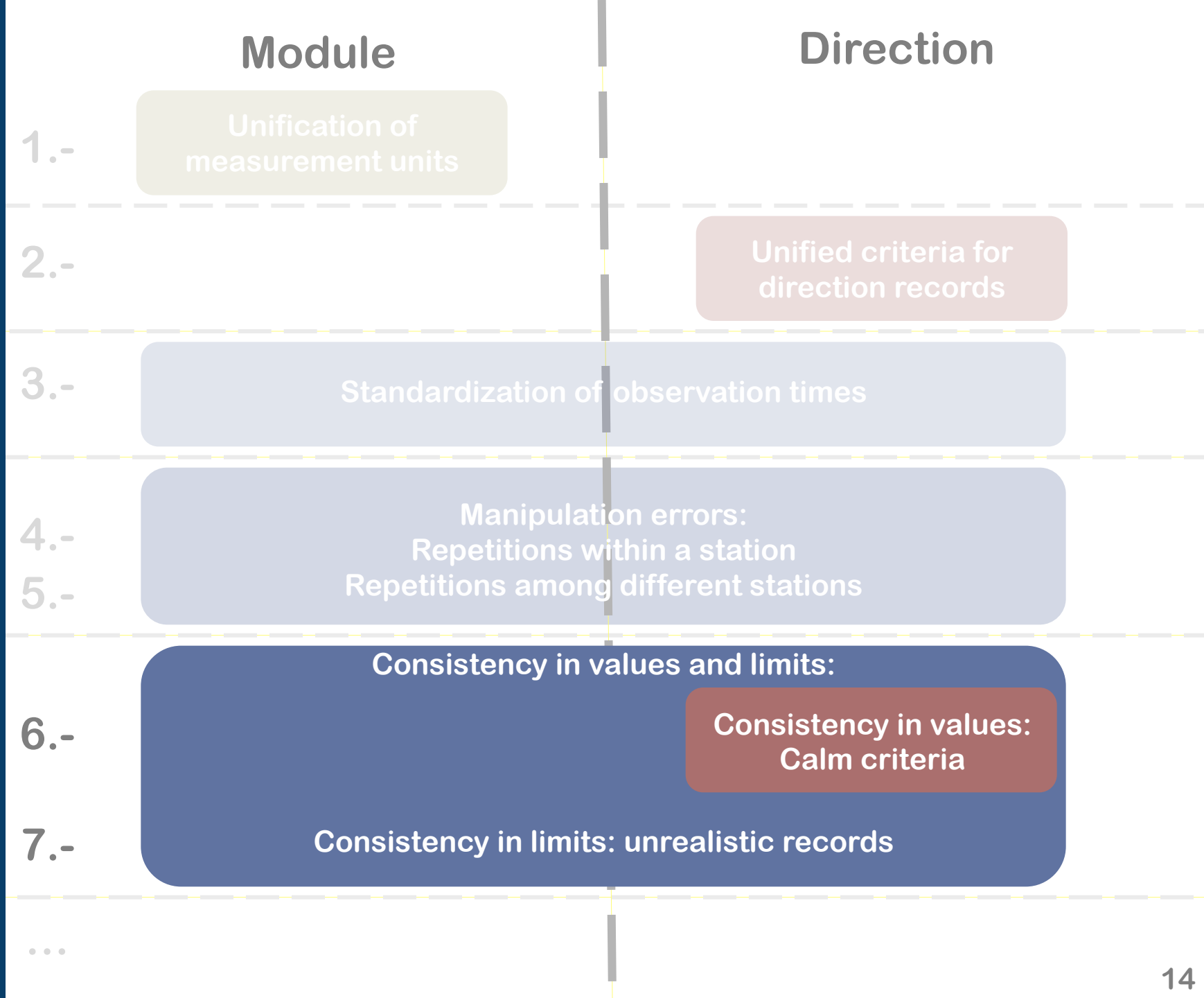


Station 495 (NCAR) shares all the data of stations 14 and 18 (EC).



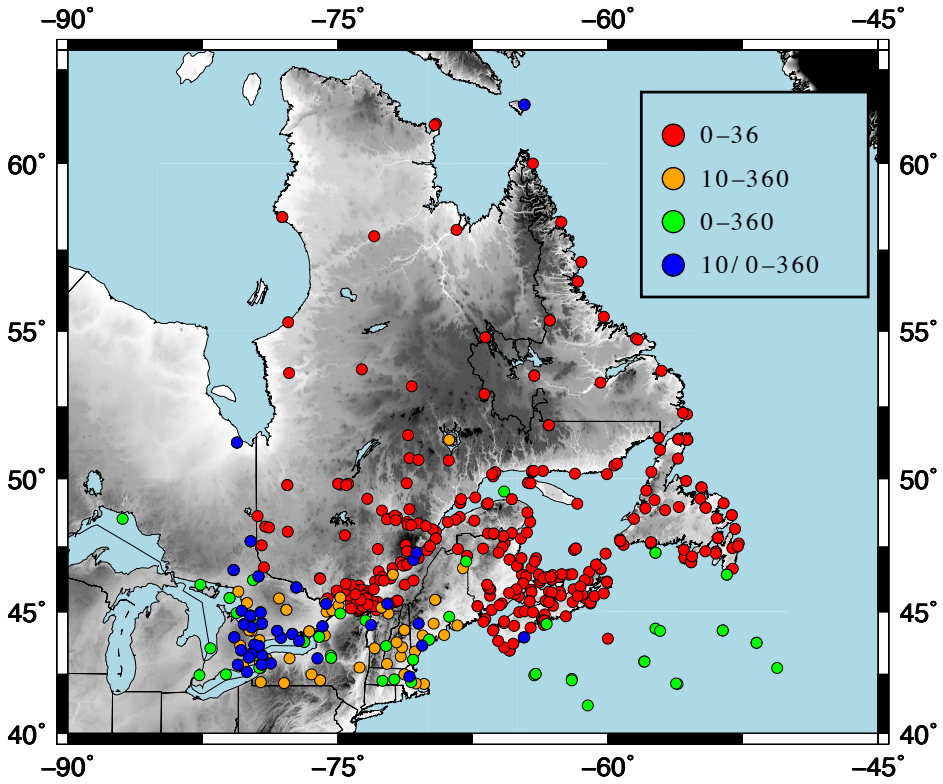
Number of days found to be repeated at various sites in the database.

Motivation
Database Compilation
Quality Assurance



Unification of calm criterion:

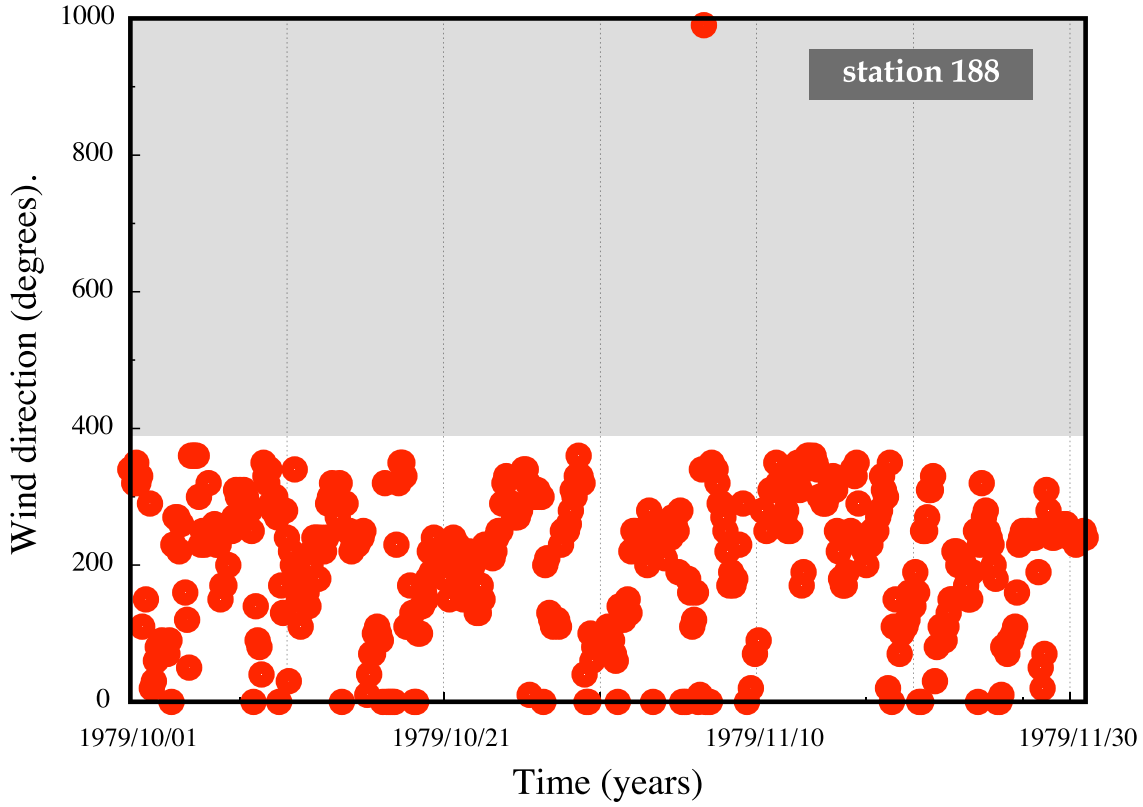
- **Calms are defined as: module=0,direction=0**
- **True North is defined as: direction=360**



Ranges of direction. All the ranges are set at [0,360].

Unrealistic records:

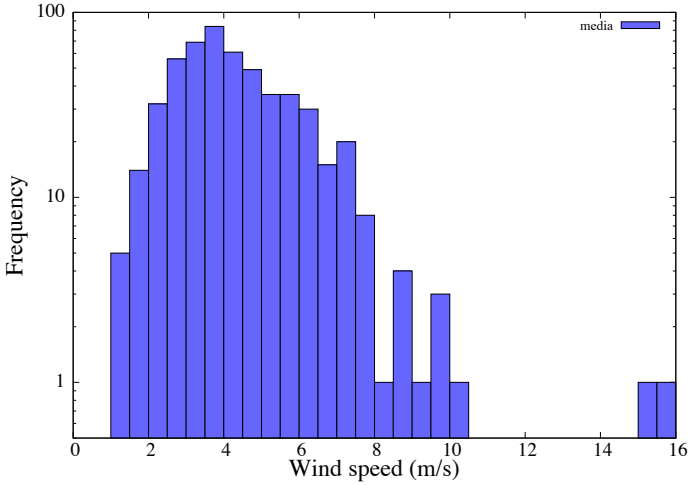
- **Direction: anything out of [0,360] is erased.**

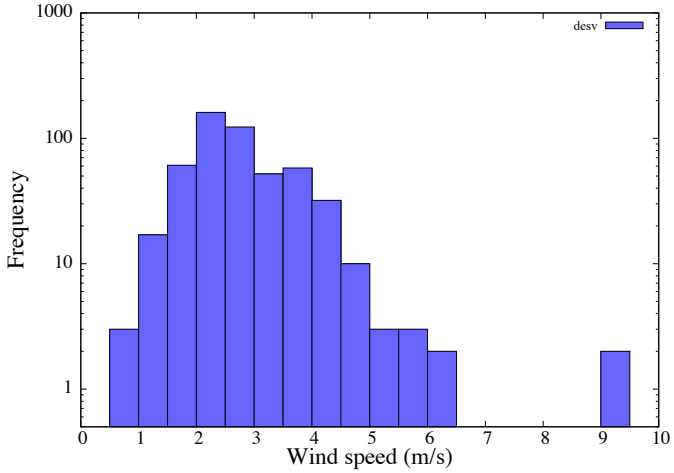


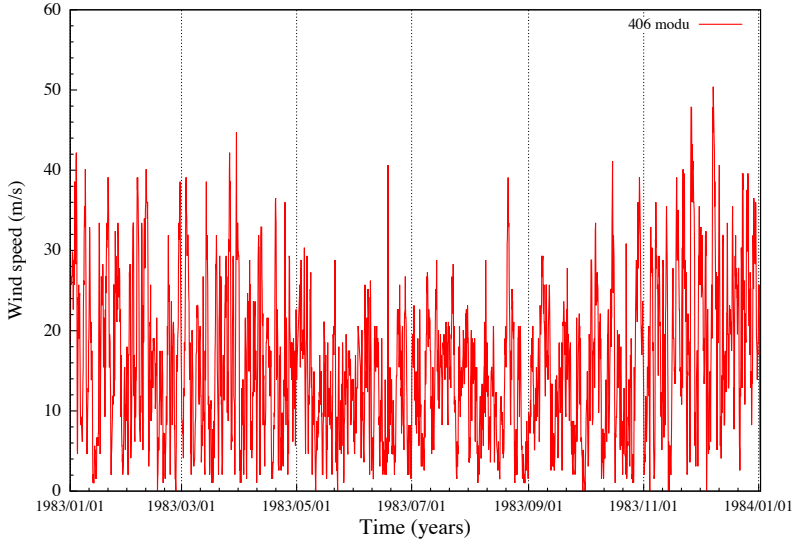
Example of physically impossible direction records.

Unrealistic records:

- **Direction: anything out of [0,360] is erased.**
- **Module: any negative value or beyond the anemometer boundary is erased.**

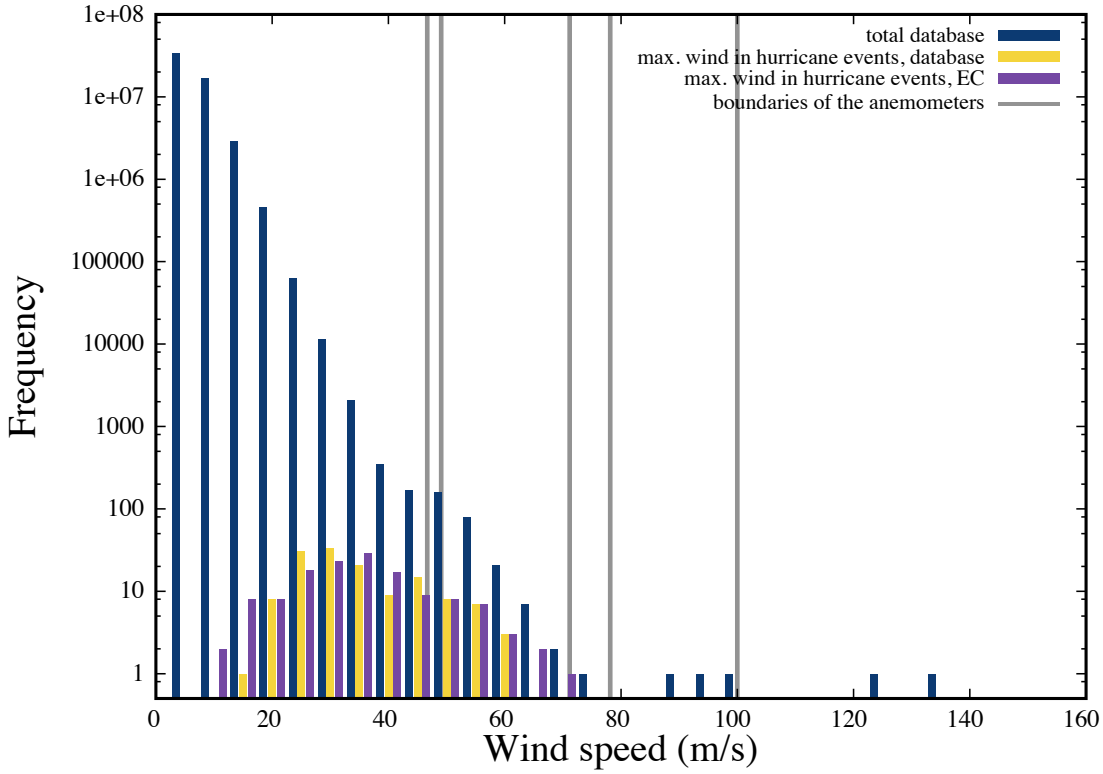






Unrealistic records: [IN PROGRESS…]

- **Direction: anything out of [0,360] is erased.**
- **Module: any negative value or beyond the anemometer boundary is erased. But we don't have that info…**
- **Alternative approach: find the maximum speeds in hurricane events to look for a reasonable threshold to begin with… (2012-V-28)**

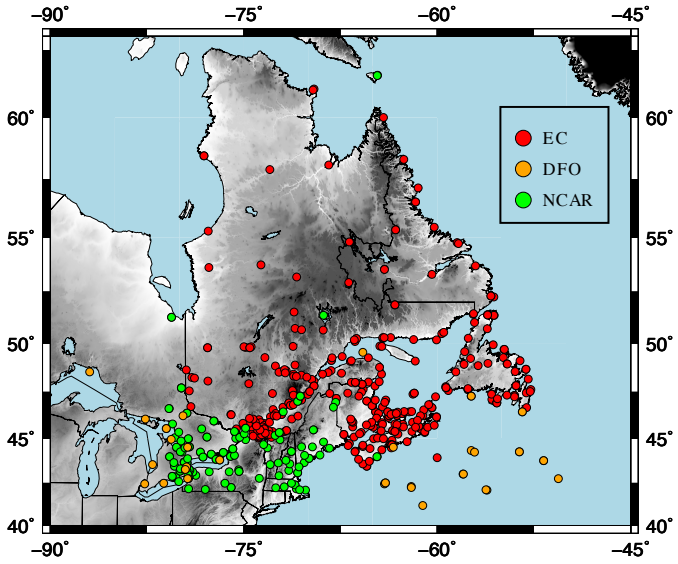


Histogram of wind speeds for all the database (blue), histogram of max. wind speeds in each hurricane event in our database (yellow), histogram of max. wind speeds from EC web (purple). Grey lines are the upper limits of the anemometers.

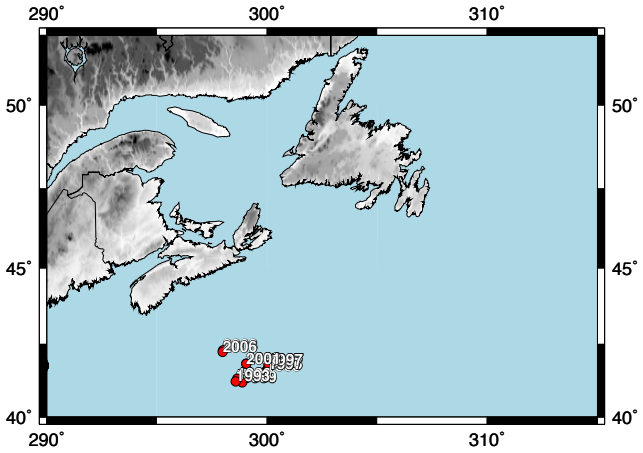
Thank you!

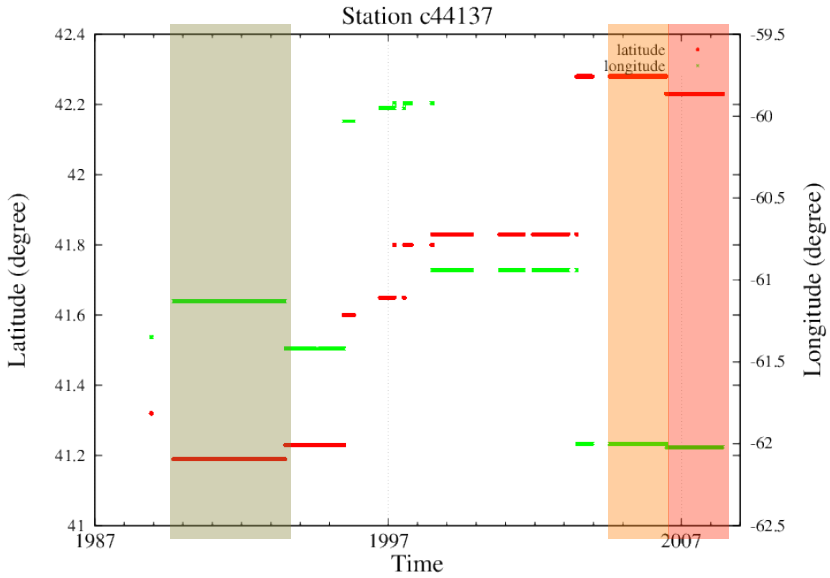
Motivation Database Complitaion Quality Assurance

Example of problems: BUOYS

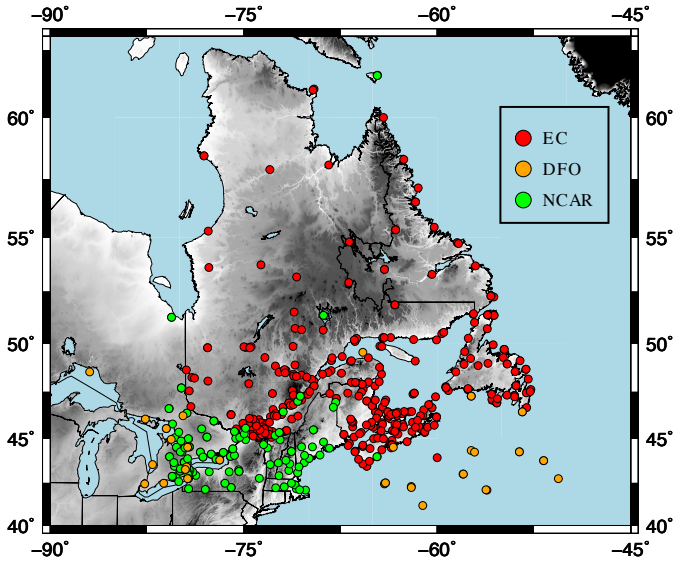


- The buoys are displaced by the institution from time to time.
- We consider each position as an independent buoy.
- We only retain buoys with at least one year of data.



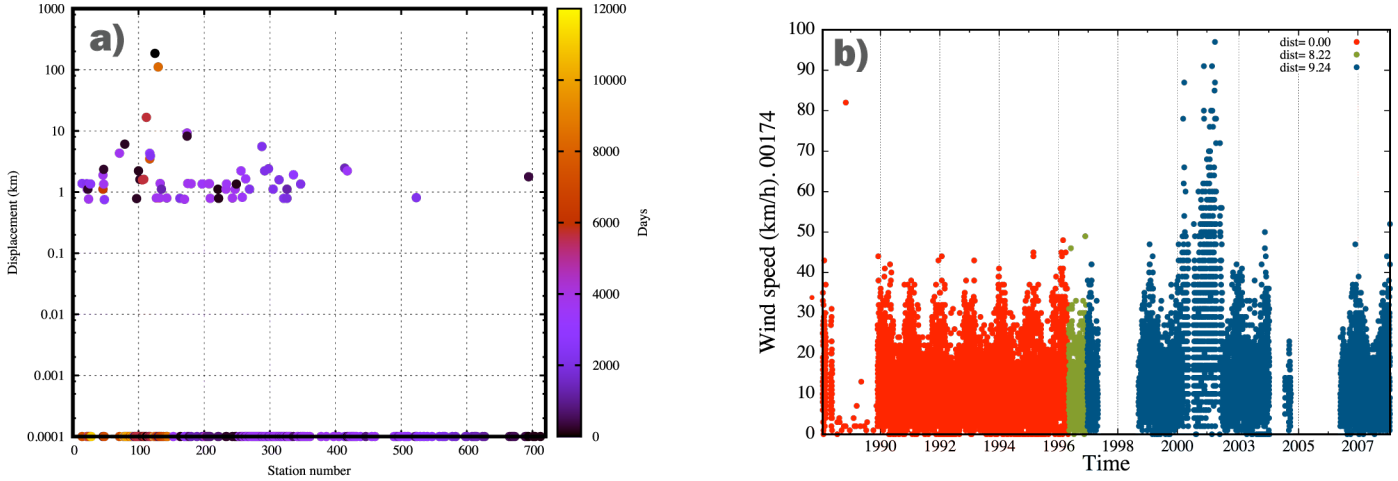


Example: buoy c44137 which belongs to the East Coast side. (a) shows the year of the first data in each location, (b) shows the latitude (red) and longitude (green) thru time, the painted periods are the only ones that were kept.

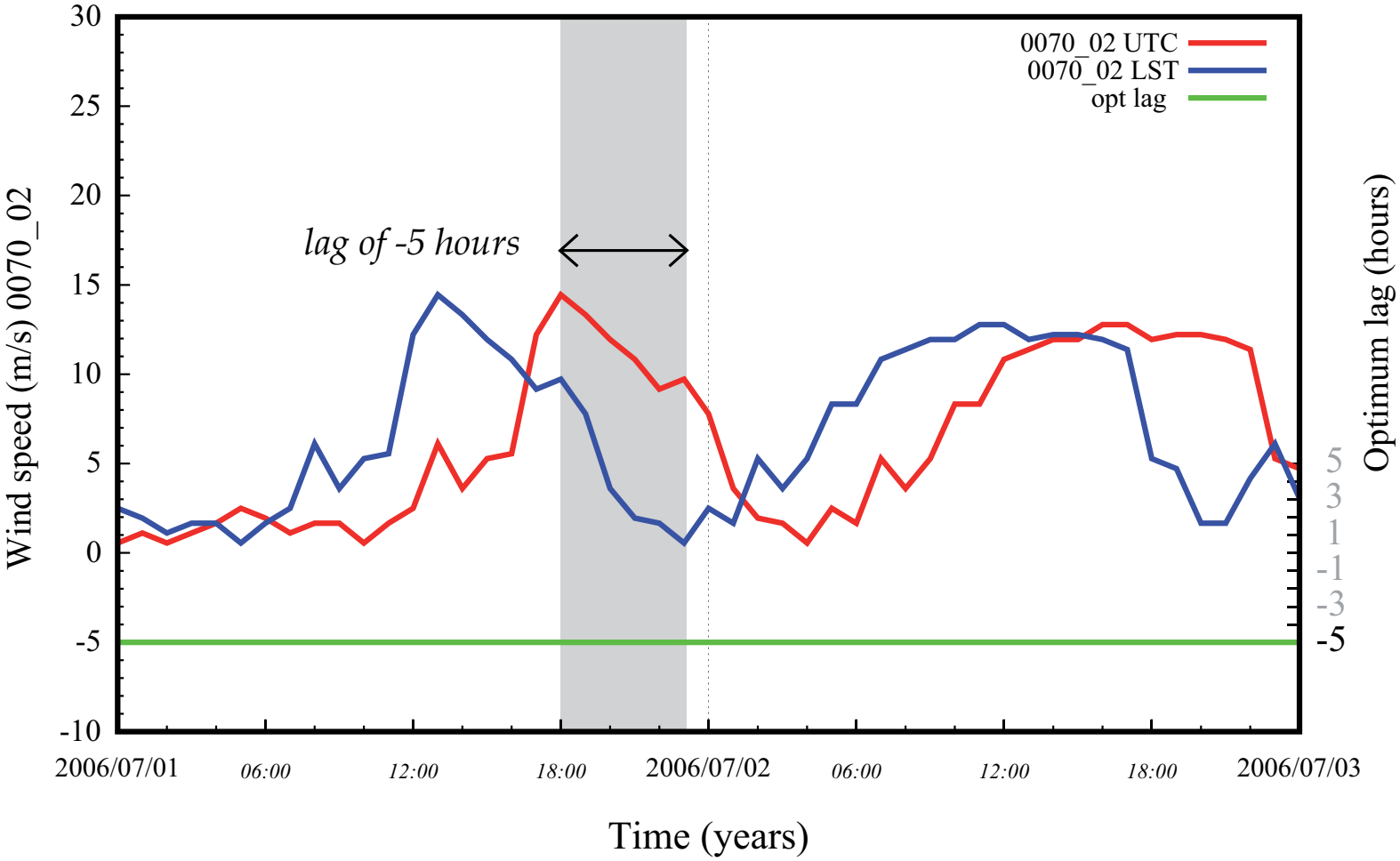


Example of problems: NCAR

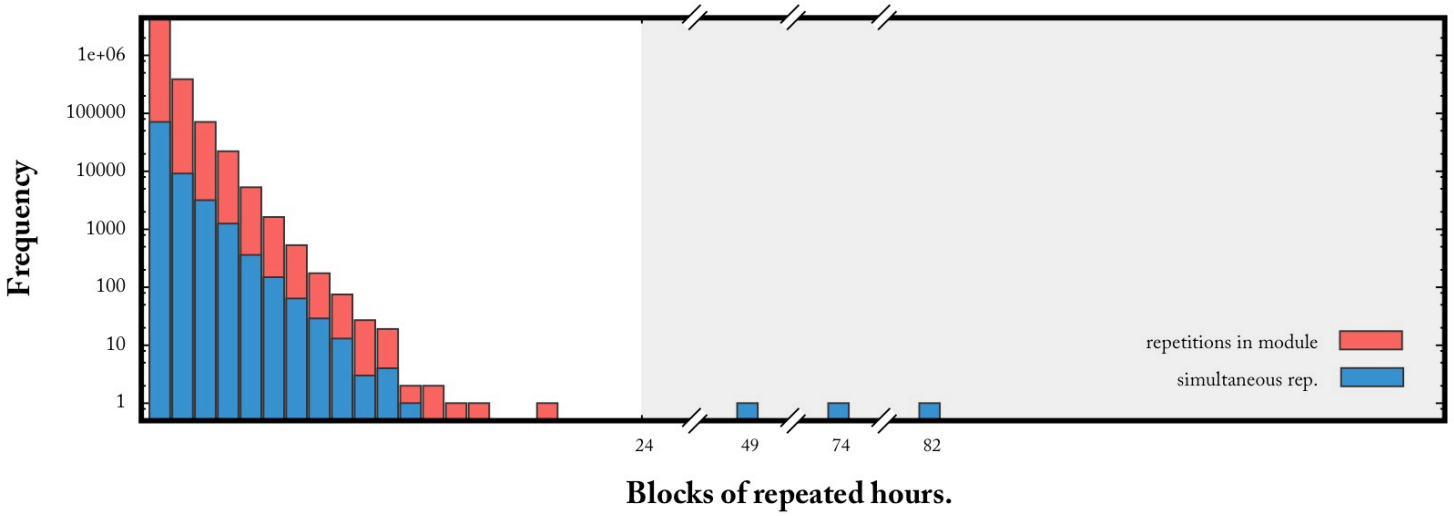
- World Meteorological Organization (WMO) provides a finite number of station identifiers (ID) to each country.
- When the countries are running out of Ids they tend to reuse the ID to a closer newer stations of the last.
- As a result, each ID shares positions and data from different real stations.



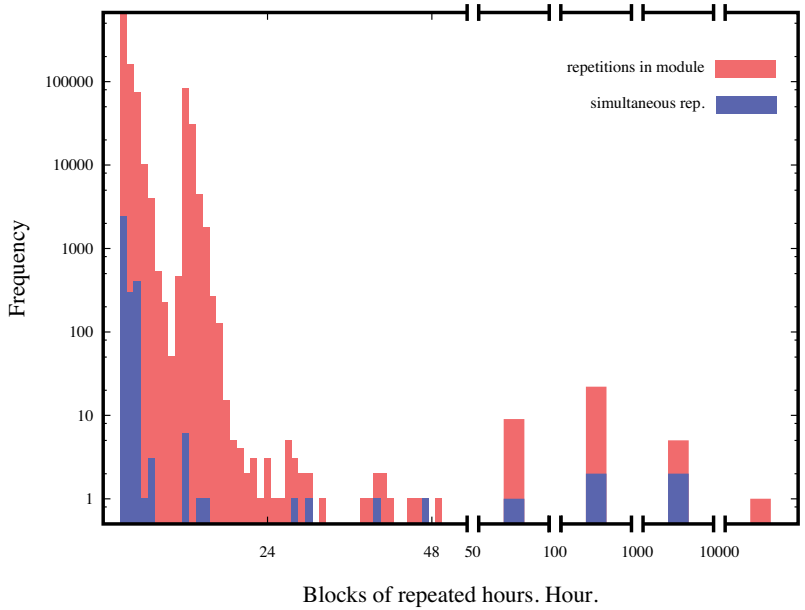
Distance and duration for ID position shifts for all the stations (a) (each station is represented in a vertical line). Example of an station with 3 shifts (b).



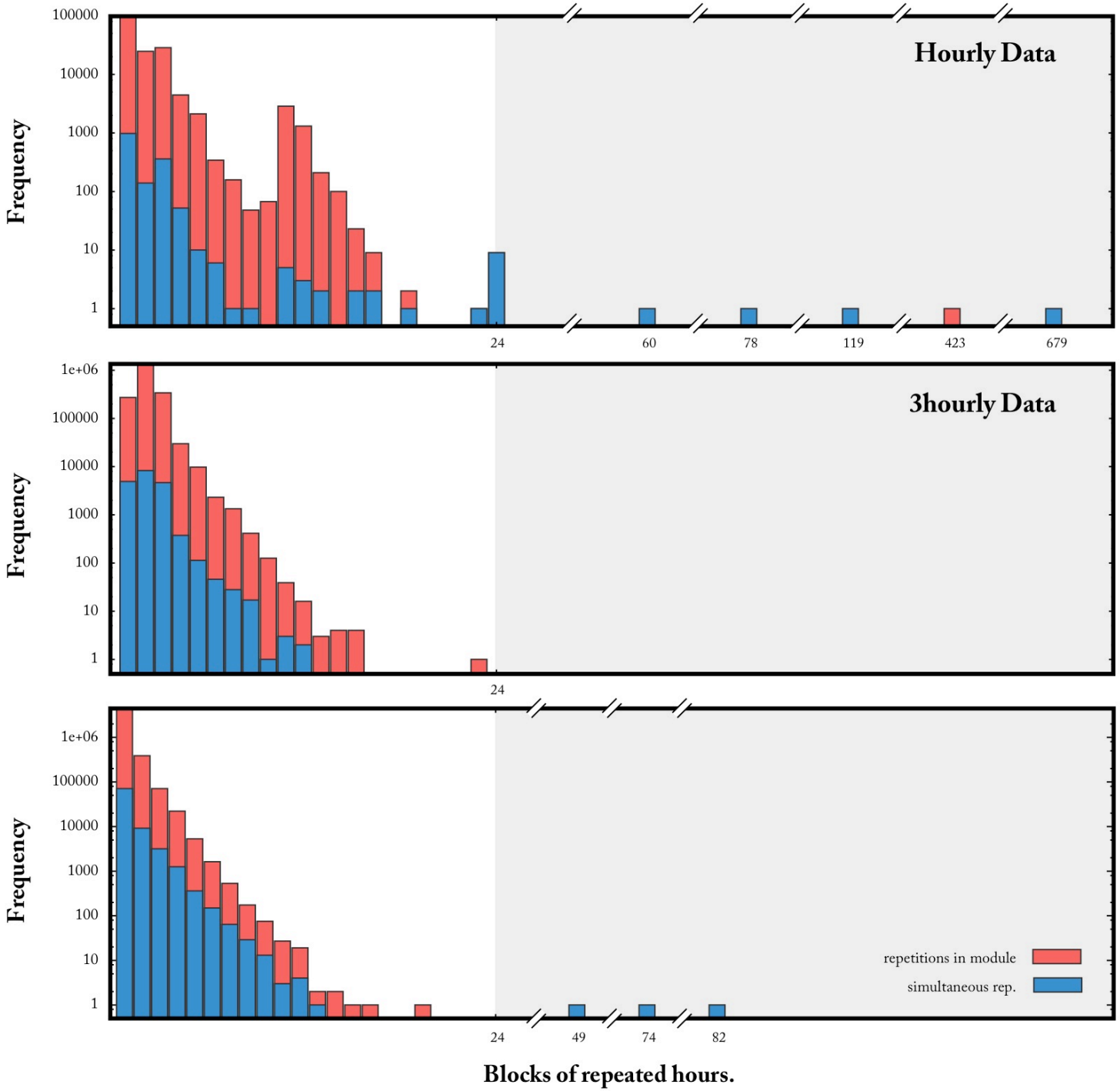
Example of the time lag between the station's LST and UTC.



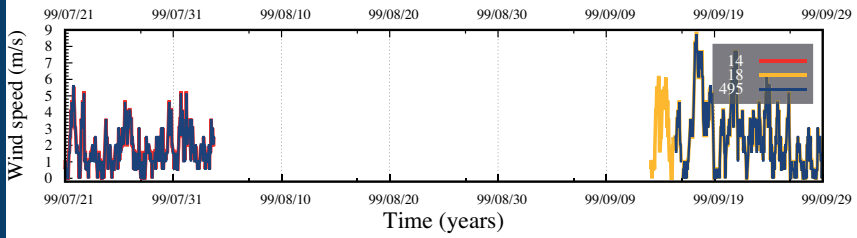
Distribution for repetitions within a station.

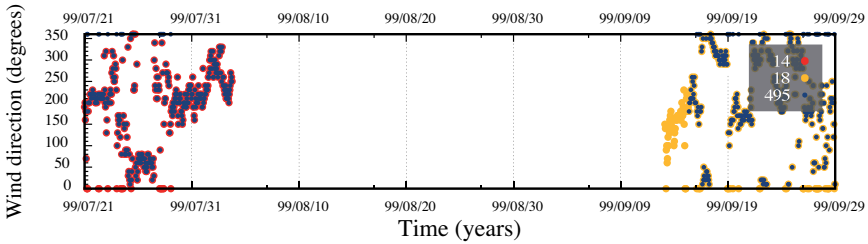


Distribution for repetitions among different stations.

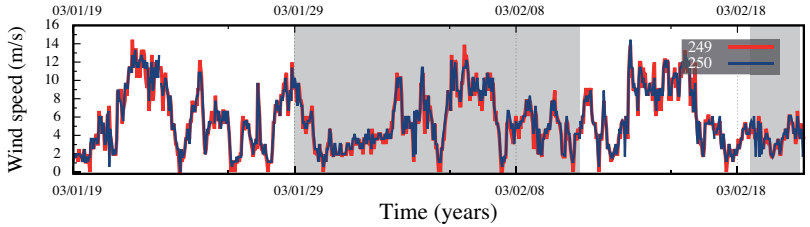

Hourly Data
Frequency
100000
10000
1000
100
10
1
24
60
78
119
423
679
3hourly Data
Frequency
1e+06
100000
10000
1000
100
10
1
24
Frequency
1e+06
100000
10000
1000
100
10
1
24
49
74
82
repetitions in module
simultaneous rep.
Blocks of repeated hours.

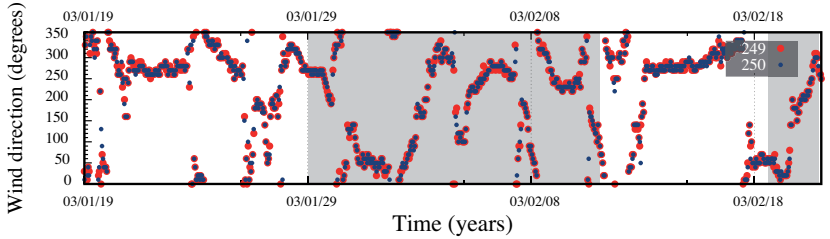
Repetitions among stations:



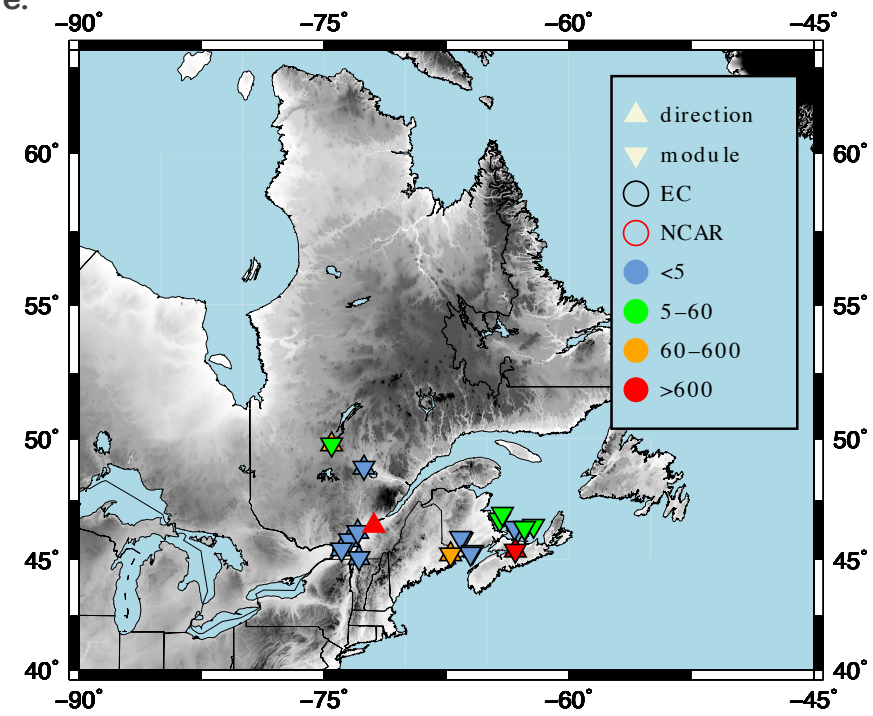


Station 495 (NCAR) shares all the data of stations 14 and 18 (EC).





Neighbour stations 249 and 250 share 10 days of data but the also have a similar behaviour. This is a false positive.



Number of days found to be repeated at various sites in the database.